The world's largest endgame study database - edition 5 (HHdbV)

Introduction

The fifth edition of the famous Harold van der Heijden endgame study database (HHdbV) is available now. The new edition, with 85,619 studies not only has more than 9,000 new studies in comparison with HHdbIV, but also the solutions of tens of thousands of studies have been corrected or updated. It is by far the most comprehensive collection of endgame studies available.

Chess players can benefit from endgame studies by trying to solve them. This trains both one's calculation ability and tactical performance in the endgame.

For the endgame study enthusiast, either admirer, cook hunter, composer or tourney judge HHdbV is indispensable. Apart from the new studies the new version has many improvements over the previous editions.

www.hhdbv.nl

What is an endgame study?

An endgame study is a chess position presented as a puzzle with the stipulation *White wins*, or *White draws*, and has a unique solution. Although it looks like a game fragment, an endgame study has been composed. A good endgame study should have an entertaining solution with surprising moves or beautiful combinations that baffle chess players. It is also fun to try and solve endgame studies for both beginners and world class grandmasters as difficulty ranges from simple to very difficult.

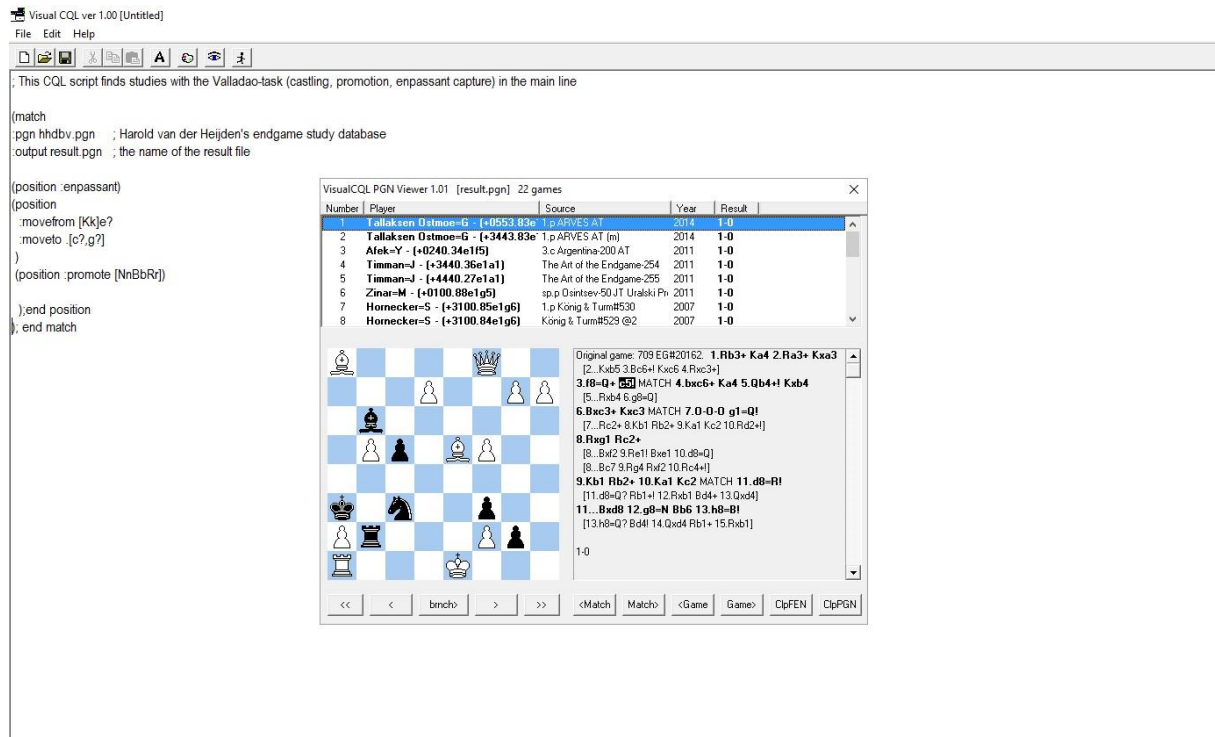http://en.wikipedia.org/wiki/Endgame_study

Software

The database is in PGN-format. This is a standard chess database format and can be accessed by commercial chess database programs (like ChessBase, Chess Assistant), commercial chess playing software (Fritz, Rybka, Shredder, etc.) and many freeware programs (e.g. PGN-readers).

The PGN-format was developed for storing chess games, but is also suitable for a database of endgame studies. Apart from the initial position and the solution (including sublines or analysis) also additional information is provided: the name(s) of the composer(s), the GBR-code which is an index code denoting the chess force in the initial position, place and date of the primary source (tourney, journal, magazine) and whether it is a win or a draw study.

The chess software mentioned above can be used to find studies in the database by name, year, source, material balance, and numerous other criteria. Most chess software also allows you to find positions or moves. For advanced users (like judges and composers) a magnificent tool was especially developed

for the endgame study database by Lewis Stiller and Gady Costeff for finding complex patterns and manoeuvres: Chess Query Language (CQL). CQL as well as the graphical interface Visual CQL (programmed by Emil Vlasák) are freely downloadable.

Screenshot of Chess Query Language and Visual CQL.

http://en.wikipedia.org/wiki/Portable_Game_Notation
http://en.wikipedia.org/wiki/ChessBase
http://en.wikipedia.org/wiki/Chess_Assistant
http://en.wikipedia.org/wiki/List_of_chess_software
http://en.wikipedia.org/wiki/GBR_code
http://rbnn.com/cql/
http://www.vlasak.biz/vcql.htm
http://www.xs4all.nl/~timkr/chess2/cql.htm

What's new in HHdbV? (information for experienced users)

HHdbV contains considerably more information than HHdbIV did. Apart from almost 9,500 new studies, also the solutions of tens of thousands of studies have been updated, e.g. by adding sublines from primary sources. Also, many endgame study enthusiasts reported thousands of cooks in studies to Harold.

Like in HHdbIV sources of corrections, modifications or versions are added as text before the 1st move of the solution. The same goes for the cooks, including the names of the person(s) who cooked the study (and the relevant move is marked with his initials). This information is relevant for people writing about endgame studies. Without HHdbV it is almost impossible to find out where a certain correction of a study was published, or who managed to first cook that famous prize winner!

Other details/additions (☑ = new in HHdbV):

- Publication year of pre-1792 studies added.
- If known, exact publication dates (e.g. newspapers) or publication month (magazines) are included.
- 100% anticipations: reference to the relevant study added.
- For studies that were inspired on a previous study ("after"): reference to the relevant study added.
- EG-number included.
- For some studies, the composer mentioned the composition date (year) of a study which (of course) was earlier than the year of publication. This information is included.
- Results of ring tourneys added ☑
- Source information added if a study was published as an original in more than one source ☑
- In some cases, orthodox chess problems are sometimes published as win studies. If known, the original stipulation is provided ☑

Similar to HHdbIV, codes are included to be able to identify cooks, corrections and other relevant properties of the endgame studies:

@1: second solution (at move 1).

@2: cook; extra solution after move 1.

@3: incorrect; White is unable to fulfil the stipulation (in a win study Black draws or wins; in a draw study White loses).

@4: "super-cook". White can even win in a study with a draw stipulation.

(@1) or (@2): minor duals.

@1?, @2?, @3?, @4?: suspicious, or claim without analysis.

(c): correction; i.e. original study was unsound.

(m): modification; i.e. original study was sound, the improvement has another motivation.

(v): version (perhaps a correction or a modification).

(s): corrected solution (without changing the position).

(ea): too many composers' names to fit. All names given as text before the first move.

(tw): twin study (also triplicates, quadruplicates, etc).

(pl): plagiarism or accidental re-composition (it does happen!). Although there is no doubt in many a case, all studies in the database are only considered "suspect".

(ph): posthumously published.

(te): theoretical ending (i.e. probably not an endgame study).

(cr): colours reversed (the original stipulation was, for instance: Black to play and win).

(ce): computer ending (EGTB-derived ending).

(tt): theme tourney.

Statistics (by Harold van der Heijden, December 2015)

The previous editions from 1991, 2000, 2005, and 2010 contained 23358, 58801, 67691, and 76132 endgame studies, respectively.

HHdbV holds 85,619 endgame studies by no less than 5,505 composers, of which a vast 4,506 have less than 10 studies in the database (and 2,404 only a single study!). All these figures include unsound studies, corrections, and versions.

The average number of studies per composer is 16.9 (which is not the number of studies divided by the number of composers, as there are co-authored studies). The top ten of the most prolific composers is almost unchanged in comparison with HHdbIV (see Table 1), although the number of studies per

composer is sometimes considerably higher because corrections were published. The late Iuri Akobia is the only newcomer in the top ten.

Table 1: The top ten of composers with the largest number of studies in HHdbV. Between brackets the number of studies without corrections or versions are given.

1. Ernest Pogosyants 2,178 (1,816)
2. Henri Rinck 1,778 (1,532)
3. Alexey Troitzky 1,750 (953)
4. Ladislav Prokes 1,247 (1,055)
5. David Gurgenidze 958 (756)
6. Michael Bent 941 (801)
7. Iuri Akobia 815 (701)
8. Bernard Horwitz 774 (614)
9. Ghenrikh Kasparyan 753 (513)
10. Filip Bondarenko 731 (601)

It is tempting to make a similar list with only the studies that are sound. But as studies of certain composers have been systematically checked using state-of-the-art hardware and software and those of others have not, such a list would have considerable bias.
Some composers made a giant leap in the number of studies (again including corrections) between HHdbIV and HHdbV: P. Arestov (+297), I. Akobia (+287), M. Garcia (+253), R. Becker (+247), P. Krug (+246), E. Pogosyants (+224), and J. Mikitovics (+208) added more than 200 studies to their oeuvre.

The numbers of studies per decade sharply increased in the 1920's and again in the 1970's (see Table 2). Since the 1970's on average 1000 new studies have been published per year. HHdbIV appeared exactly 5 years ago, and the database has grown with almost 10,000 studies. Apart from old studies discovered, this must be largely due to new corrections of older studies. This means that if we keep the same pace, the 100,000 mark will be crossed in 2023!

Table 2: The total number of studies per decade in HHdbV.

1881-1890	612
1891-1900	908
1901-1910	1,880
1911-1920	2,212
1921-1930	6,781
1931-1940	6,135
1941-1950	5,230
1951-1960	7,127
1961-1970	7,784
1971-1980	10,584
1981-1990	10,730
1991-2000	9,218
2001-2010	9,579
2011-2015	(4,117)

Some people wonder why I do not "simply check" all the studies in my database by computer when I include them in my database, or do that in retrospect. Apart from the work involved (it would take several years just to check every study at a rate of one per 5 minutes and working on this for several hours per day) I would have to repeat this with every generation of hardware, software and EGTB. This would leave me no time to add new studies to my database. No less than 35% of the studies in HHdbV seem to be cooked (Table 3). It must be noted that it is hardly an exception that, with contemporary facilities, also claims/cooks are refuted, making the study sound again. I have decided to continue to include unsoundness claims in my database, despite the fact that almost anybody is able to check the soundness of endgames studies today. But, especially for organizers for solving events, I underline that studies which are not reported to be unsound in my database, should be thoroughly checked before using them in your event. Of course, if you do find a problem in a candidate study, be sure to let me know.

Table 3: Statistics on cooks and corrections in HHdbV

Total	85,619
@1	7,525
@2	15,559
@3	11,078
@4	336
Sound	55,794 (65.1%)
(c)	6,348
(m)	854
(v)	2,696

In total, HHdbV has 3,807,322 moves in the main line and sublines, which is an average of 44.5 per endgame study. I have added all these moves "manually" into my database myself. As I have a strict convention (e.g. a white try, except in case of a stalemate, should always end with a black move, and there are numerous other rules) for entering moves, I never rely on PGN files supplied by others. In such cases I do not import the file into my database, but enter the moves manually playing through the PGN file in a separate screen.

Regarding cooks, of course, the @1 (cook/dual on the first move) and the @2 (cook/dual not on the first move) are sometimes somewhat subjective even if is clear that the alternative move also meets the stipulation. In particular, in QBQpp or QSQpp endings it is often difficult to see whether EGTB indicated alternative winning moves are minor duals (waste of time) or genuine alternative winning moves. There seems to be some progress in algorithms to distinguish real cooks from time wasting duals. Of course, such problems do not occur for @3 cooks (study is unsound), assuming that the claim stands up to contemporary hardware/software/EGTB.

Table 4: Most frequent moves of white and black pieces in HHdbV.

wK	Kc6	14,066	bK	Ke5	16,552
wQ	Qe5	5,725	bQ	Qe5	3,650
wR	Rb7	7,047	bR	Rb8	6,136
wB	Bd5	8,363	bB	Bd5	6,129
wS	Sf6	10,677	bS	Se5	5,588
wP	b7	9,460	bP	h2	9,423

One would guess that king moves are the most frequently played moves in endgame studies as every endgame study at least have a pair of kings. And that is correct with …Ke5 being the most frequent move. To me it comes as a surprise that Kc6 is the most frequent white move.

Personal profile

Dr. Harold van der Heijden (b. 1960) is one of the world's leading experts on endgame studies. He is: chief editor of the famous international magazine EG, FIDE master of chess composition (endgame studies), FIDE judge for endgame studies, Spokesman of the endgame study subcommittee of the WFCC, author of two books about endgame studies, and collector of books with endgame studies. But chess is only his hobby of course. Professionally, after obtaining a PhD in 2009 (Utrecht University), he became head of the R&D-laboratory of a veterinary laboratory in 2010. He is married and has two sons (26 and 23).